

## KLASIFIKASI PERUBAHAN HARGA OBLIGASI KORPORASI DI INDONESIA MENGUNAKAN METODE *NAIVE BAYES CLASSIFICATION*

Khotimatus Sholihah, Di Asih I Maruddani<sup>2</sup>, Abdul Hoyyi<sup>3</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

<sup>2,3</sup>Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro

e-mail [skhotimatus@gmail.com](mailto:skhotimatus@gmail.com)

### ABSTRACT

Bond is a medium-long term debt securities which can be sold and contains a pledge from the issuer to pay interest for a certain period and repayment of the principal debt at a specified time to the bonds buyer. Bonds price changes any time, it could be beneficial or give disadvantage to investors. Investors should know the best conditions to buy bonds on a discount, or sell them at a premium price. By classify the changing of bonds price, it could help investors to gain optimum return. One method is Naive Bayes classification. In theory, It has the minimum error rate in comparison to all other classifiers. Bayes is a simple probabilistic-based prediction technique which based on the application of Bayes theorem with strong independence assumptions. Before classifying, preprocessing data is required as a stage feature selection. In this case, the Mann Whitney test can be done to choose the independent features of each class. Validation technique in use is k-fold cross validation. Based on analysis, we gained average accuracy at 78,52% and 21,8% error. With high accuracy and quite low error, it means that the Naïve Bayes method works quite well on classifying the corporate bonds price changes in Indonesia.

**Keywords:** bonds, classification, k-fold cross validation, Naive Bayes

### 1. PENDAHULUAN

Pasar modal merupakan sarana pembentuk modal dan akumulasi dana yang diarahkan, untuk meningkatkan partisipasi masyarakat dalam pengarahana dana guna menunjang pembiayaan pembangunan nasional (Ahmad, 2003). Di Indonesia terdapat berbagai macam sekuritas yang diperjualbelikan di dalam pasar modal, salah satunya adalah obligasi. Obligasi adalah surat utang jangka menengah-panjang yang dapat dipindahtangankan yang berisi janji dari pihak yang menerbitkan untuk membayar imbalan berupa bunga pada periode tertentu dan melunasi pokok utang pada waktu yang telah ditentukan kepada pihak pembeli obligasi tersebut (Wijayanto, 2015).

Seperti halnya instrumen investasi lainnya, obligasi juga mengalami perubahan dalam harga. Ada banyak faktor yang mempengaruhi perubahan harga obligasi. Fabozzi (2000) menyebutkan bahwa faktor yang mempengaruhi perubahan harga obligasi adalah kupon, jangka waktu jatuh tempo dan tingkat hasil di mana obligasi diperdagangkan (*yield*). Menurut Anoraga dan Pakarti (2001), obligasi ada yang disebut obligasi dengan premium (*at premium bond*) yaitu kondisi di mana harga pasar obligasi lebih tinggi di atas nilai nominalnya, ada pula yang disebut obligasi dengan diskon (*at discount bond*), yaitu kondisi di mana harga pasar obligasi berada di bawah nilai nominalnya. Dengan mengklasifikasikan perubahan harga obligasi dapat membantu investor untuk mendapatkan *return* yang optimal.

Ada banyak metode klasifikasi, antara lain *Naive Bayes*. Metode *Naive Bayes* merupakan metode yang mudah diimplementasikan. Kelebihan lain metode ini adalah dapat menangani *missing value*, dan apabila asumsi saling bebas terpenuhi, maka tingkat akurasi sangat tinggi (Sibarani, 2008). Berdasarkan uraian tersebut, penulis tertarik untuk mengklasifikasikan perubahan harga obligasi menggunakan metode *Naive Bayes Classification*. Data yang digunakan adalah data obligasi korporasi di Indonesia tanggal 30 November 2015 s.d.1 Desember 2015 dengan tipe obligasi *fixed rate bond*.

## 2. TINJAUAN PUSTAKA

### 2.1. Pengertian Obligasi

Zubir (2012) menyatakan bahwa obligasi adalah surat utang yang diterbitkan oleh pemerintah atau perusahaan untuk mendapatkan dana. Pihak penerbit obligasi tersebut akan membayarkan sejumlah bunga atau sering juga disebut sebagai *coupon*, misalnya setiap enam bulan, kemudian pada tanggal jatuh tempo (*maturity*) utang pokoknya akan dikembalikan kepada pemegang obligasi tersebut. Menurut Fabozzi (2000) dalam Ibrahim (2008) menyatakan bahwa untuk menghasilkan suatu strategi investasi obligasi yang baik diperlukan suatu analisis terhadap faktor-faktor yang berpengaruh terhadap harga obligasi. Harga obligasi dipengaruhi oleh *risk* (risiko) dan *return* (hasil) yang diharapkan dari obligasi.

### 2.2. Karakteristik Obligasi

Krisnilasari (2007) menyatakan bahwa adapun karakteristik umum yang tercantum pada sebuah obligasi hampir mirip dengan karakteristik pinjaman utang pada umumnya yaitu meliputi nilai penerbitan obligasi, jangka waktu obligasi, tingkat suku bunga obligasi (*coupon*), jadwal pembayaran suku bunga dan jaminan.

### 2.3. Peringkat Obligasi Perusahaan

Zubir (2012) menyatakan bahwa simbol peringkat obligasi perusahaan menggambarkan kondisi kesehatan perusahaan sebagai berikut:

**Tabel 1.** Peringkat Obligasi Perusahaan

Peringkat	Keterangan
AAA	Obligasi yang mempunyai risiko investasi yang sangat kecil. Pembayaran bunga dan pokok pinjamannya dijamin oleh laba operasi yang besar dan stabil.
AA	Obligasi dengan kualitas baik yang dapat memenuhi kewajiban bunga dan pokok pinjamannya.
A	Obligasi dengan jaminan terhadap bunga dan utang pokok cukup memadai tetapi dikhawatirkan jaminan tersebut dapat menyimpang di masa yang akan datang.
BBB	Dimana keamanan pembayaran bunga dan pokok pinjamannya dinilai cukup saat ini, tetapi jaminan tertentu kurang atau tidak reliabel dalam waktu yang lebih panjang.
BB	Pembayaran bunga dan pokok pinjaman sering kurang aman di masa depan.
B	Obligasi ini kurang memenuhi karakteristik investasi karena rendahnya jaminan.
CCC	Obligasi yang tingkat spekulasinya tinggi.
CC	Obligasi dengan tingkat spekulasi yang sangat tinggi. Obligasi tersebut sering gagal bayar.
C	Obligasi yang sangat spekulatif dan dianggap mempunyai prospek yang sangat buruk.
D	Obligasi dengan kelas paling rendah yang menyatakan bahwa perusahaan tersebut gagal bayar.

### 2.4. Konsep Klasifikasi

Menurut Han dan Kamber (2006), klasifikasi adalah proses pencarian sekumpulan model atau fungsi yang menggambarkan dan membedakan kelas data dengan tujuan agar

model tersebut dapat dipergunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya atau dapat memprediksi kecenderungan data-data yang muncul di masa depan. Sementara menurut Prasetyo (2014), klasifikasi dapat didefinisikan secara detail sebagai suatu pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target  $f$  yang memetakan setiap vektor (*set fitur*)  $x$  ke dalam satu dari sejumlah label kelas  $y$  yang tersedia. Pekerjaan pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori.

## 2.5. Probabilitas

Subanar (2013) menyebutkan bahwa seandainya bila kejadian  $Y$  terjadi, maka agar  $X$  terjadi, kejadian sebenarnya haruslah berupa titik-titik yang berada dalam  $X$  dan  $Y$ , yaitu harus berada dalam  $X \cap Y$ . Karena  $Y$  telah terjadi, maka  $Y$  menjadi ruang sampel baru sehingga probabilitas kejadian  $X \cap Y$  terjadi sama dengan probabilitas  $X \cap Y$  relatif dengan probabilitas  $Y$ .

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \quad (1)$$

$P(X|Y)$  hanya terdefinisi bila  $P(Y) > 0$ . Dari persamaan tersebut dapat diperoleh:

$$P(X \cap Y) = P(X|Y)P(Y) \quad (2)$$

Misalkan  $X$  dan  $Y$  kejadian, dan  $X$  dapat dinyatakan sebagai berikut:

$$X = (X \cap Y) \cup (X \cap Y^c)$$

Karena  $(X \cap Y)$  dan  $(X \cap Y^c)$  saling asing, maka dapat dinyatakan

$$\begin{aligned} P(X) &= P(X \cap Y) \cup P(X \cap Y^c) \\ &= P(X|Y)P(Y) + P(X|Y^c)P(Y^c) \\ &= P(X|Y)P(Y) + P(X|Y^c)(1 - P(Y)) \end{aligned} \quad (3)$$

Misalkan  $Y_1, Y_2, \dots, Y_n$  adalah kejadian-kejadian yang saling asing sedemikian hingga  $\bigcup_{i=1}^n Y_i = \Omega$ . Misalkan  $X$  kejadian dalam  $\Omega$ . Maka

$$\begin{aligned} X &= (X \cap Y_1) \cup (X \cap Y_2) \dots \cup (X \cap Y_n) \\ &= \bigcup_{i=1}^n (X \cap Y_i) \end{aligned} \quad (4)$$

Dengan menggunakan kenyataan kejadian-kejadian  $X \cap Y_i$ ,  $i = 1, 2, \dots, n$  saling asing, maka

$$\begin{aligned} P(X) &= \sum_{i=1}^n P(X \cap Y_i) \\ &= \sum_{i=1}^n P(X|Y_i)P(Y_i) \end{aligned} \quad (5)$$

Persamaan (5) menunjukkan bagaimana untuk kejadian-kejadian  $Y_1, Y_2, \dots, Y_n$  dengan satu dan hanya satu harus terjadi,  $P(X)$  dapat dihitung dengan mensyaratkan pada salah satu  $Y_i$  yang terjadi. Misalkan  $X$  telah terjadi, untuk menentukan mana salah satu dari  $Y_j$  yang terjadi, maka

$$\begin{aligned} P(Y_j|X) &= \frac{P(X \cap Y_j)}{P(X)} \\ &= \frac{P(X|Y_j)P(Y_j)}{\sum_{i=1}^n P(X|Y_i)P(Y_i)} \end{aligned} \quad (6)$$

## 2.6. Naive Bayes Classification

*Naive Bayes* merupakan salah satu metode klasifikasi dalam data mining. *Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (aturan Bayes) dengan asumsi independensi yang kuat (naif). Dengan kata lain, dalam *Naive Bayes* model yang digunakan adalah “model fitur independen” (Prasetyo, 2013). Sibaroni (2008) menyatakan bahwa meskipun asumsi independensi cukup sulit dipenuhi dalam dunia nyata, akan tetapi dalam praktiknya masih berfungsi

dengan baik. Selain itu, menurut Han dan Kamber (2006) klasifikasi *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan jumlah yang besar.

Menurut Prasetyo (2013), jika  $X$  adalah vektor masukan yang berisi fitur dan  $Y$  adalah label kelas, *Naive Bayes* dituliskan dengan  $P(Y|X)$ . Notasi tersebut berarti probabilitas label kelas  $Y$  didapatkan setelah fitur-fitur  $X$  diamati. Notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk  $Y$ , sedangkan  $P(Y)$  disebut probabilitas awal (*prior probability*).

Selama proses pelatihan harus dilakukan pembelajaran probabilitas akhir  $P(Y|X)$  pada model untuk setiap kombinasi  $X$  dan  $Y$  berdasarkan informasi yang didapat dari data latih. Dengan membangun model tersebut, suatu data uji  $X$  dapat diklasifikasikan dengan mencari nilai  $Y$  dengan memaksimalkan nilai  $P(Y|X)$  yang didapat. Formula *Naive Bayes* untuk klasifikasi adalah

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (7)$$

Dengan  $P(X)$  adalah probabilitas  $X$  yang nilainya selalu tetap

$P(Y|X)$  adalah probabilitas data dengan vektor  $X$  pada kelas  $Y$

$P(Y)$  adalah probabilitas awal kelas  $Y$

$\prod_{i=1}^q P(X_i|Y)$  adalah probabilitas independen kelas  $Y$  dari semua fitur dalam vektor  $X$ .

Menurut Santoso (2007), persamaan (7) dapat ditulis sebagai berikut :

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Jadi dengan mengetahui nilai  $X$ , maka probabilitas prior  $P(Y)$  dapat diubah menjadi probabilitas posterior  $P(Y|X)$  yaitu probabilitas keluarnya hasil  $Y_j$  jika diketahui nilai  $X$  tertentu. Perkalian *likelihood* dengan *prior* adalah hal paling penting untuk menemukan *posterior*. Karena *evidence* bisa dianggap sebagai faktor skala sehingga hasil penjumlahan probabilitas posterior sama dengan satu. Han dan Kamber (2006) juga menyebutkan bahwa nilai  $P(X)$  adalah konstan untuk semua kelas, sehingga hanya  $P(X|Y_j) P(Y)$  yang dibutuhkan nilai maksimumnya.

## 2.7. Pemilihan Fitur Berbasis Statistik

Langkah yang paling sederhana dalam memilih fitur adalah mengamati setiap fitur yang dibangkitkan secara independen dan menguji kemampuan diskriminasinya pada setiap kelas. Meskipun langkah tersebut kurang efektif, akan tetapi dapat membantu untuk membuang fitur yang tidak sesuai (Prasetyo, 2014).

$H_0$  : Nilai fitur tidak berbeda secara signifikan

$H_1$  : Nilai fitur berbeda secara signifikan

Jika hipotesis nol benar, maka fitur akan dibuang/tidak digunakan. Jika hipotesis alternatif yang benar, maka fitur akan dipilih/digunakan (Prasetyo, 2014). Uji *Mann-Whitney* dapat digunakan untuk pengujian independensi. Ide dasarnya adalah menguji apakah nilai rata-rata fitur berbeda dalam dua kelas secara signifikan.

Menurut Daniel (1989), data dalam uji *Mann-Whitney* terdiri dari dua sampel random, Misalkan  $X_1, X_2, \dots, X_n$  notasi dari sampel random berukuran  $n$  dari populasi 1, dan misalkan  $Y_1, Y_2, \dots, Y_m$  notasi dari sampel random berukuran  $m$  dari populasi 2.

Hipotesis  $H_0 : E(X_1) = E(X_2)$

$H_1 : E(X_1) \neq E(X_2)$

Statistik uji:

Diberikan peringkat dari 1 sampai  $n+m$  pada dua sampel random tersebut. Misalkan  $N = n + m$  dan misalkan  $S$  adalah jumlah peringkat dari populasi 1. Maka didefinisikan statistik  $T$  sebagai berikut:

$$T = S - \frac{n(n+1)}{2}$$

Apabila  $n$  dan  $m$  lebih besar dari 20, maka digunakan pendekatan distribusi normal. Misalkan  $t$  adalah banyaknya angka yang sama untuk suatu peringkat, sehingga diperoleh rumus

$$Z_{hitung} = \frac{T - (n \cdot m / 2)}{\sqrt{\frac{nm(N+1)}{12} - \frac{nm(\sum t^3 - \sum t)}{12N(N-1)}}} \quad (8)$$

Apabila  $Z_{hitung}$  kurang dari  $Z_{(\alpha/2)}$  atau  $Z_{hitung}$  lebih dari  $Z_{(1-\alpha/2)}$  maka  $H_0$  ditolak.

## 2.8. Teknik Validasi Model

*K-fold cross validation* adalah salah satu teknik untuk memperkirakan kinerja klasifikasi. Pendekatan ini memecah set data menjadi  $k$  bagian set data dengan ukuran yang sama. Setiap kali berjalan, satu pecahan berperan sebagai set data uji sedangkan pecahan lainnya menjadi set data latih. Prosedur tersebut dilakukan sebanyak  $k$  kali sehingga setiap data berkesempatan menjadi data uji tepat satu kali dan menjadi data latih sebanyak  $k-1$  kali (Prasetyo, 2014).

## 3. METODE PENELITIAN

### 3.1. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data obligasi di Indonesia tanggal 30 November 2015 sampai dengan 1 Desember 2015 yang diperoleh dari PT. Penilai Harga Efek Indonesia dan [www.ksei.co.id](http://www.ksei.co.id).

### 3.2. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari variabel dependen yaitu perubahan harga obligasi yang dikategorikan menjadi dua, yaitu :

$Y = 0$ , kondisi harga obligasi kurang dari nilai nominal (diskon)

$Y = 1$ , kondisi harga obligasi lebih dari nilai nominal (premium)

Sedangkan untuk variabel independen, terdiri dari lima komponen, yaitu:

1. Kupon obligasi, diukur dengan menggunakan skala rasio.
2. Jangka waktu obligasi, dikategorikan menjadi tiga yaitu:
  - 1, yaitu ketika jangka waktu obligasi kurang dari 1 tahun.
  - 2, yaitu ketika jangka waktu obligasi antara 1 s.d. 10 tahun.
  - 3, yaitu ketika jangka waktu obligasi lebih dari 10 tahun.
3. Peringkat obligasi, dikategorikan menjadi empat yaitu:
 

AAA = 1	AA = 2	A = 3	BBB = 4
---------	--------	-------	---------
4. Nilai nominal, yaitu nilai jumlah dana yang dibutuhkan oleh penerbit obligasi.
5. *Yield*, yaitu hasil yang akan diperoleh investor apabila menempatkan dananya untuk dibelikan obligasi.

### 3.3. Tahapan Analisis

Adapun metode analisis yang digunakan adalah sebagai berikut :

1. Setelah data dikumpulkan, langkah selanjutnya adalah melakukan *checking data* dan *preprocessing data*. *Preprocessing data* merupakan tahapan pemilihan fitur yang sesuai dengan metode yang akan digunakan. Dalam hal ini dilakukan uji *Mann-Whitney* untuk memilih fitur yang independen.
2. Membagi data tersebut menjadi data *training* (data latih) dan *testing* (data uji) menggunakan *k-fold cross validation*.
3. Menghitung probabilitas prior ( $P(Y)$ ) dari data *testing* berdasarkan data *training*.
4. Menghitung probabilitas atribut terhadap masing-masing kelas ( $P(X_i|Y)$ ) pada data *testing* berdasarkan data *training*.
5. Menghitung perkalian probabilitas prior dengan probabilitas atribut pada masing-masing kelas ( $P(Y)P(X_i|Y)$ ).
6. Mencari nilai maksimal dari  $\frac{P(Y)P(X_i|Y)}{P(X)}$  pada kedua kelas. Nilai terbesar dari perhitungan tersebut merupakan hasil prediksi.
7. Pengukuran kinerja klasifikasi dengan menghitung nilai akurasi dan tingkat kesalahan (*error*).

## 4. HASIL DAN PEMBAHASAN

### 4.1. Pemilihan Fitur Menggunakan Uji Mann-Whitney

Prasetyo (2014) menyebutkan bahwa salah satu fase penting dalam pemrosesan awal data mining adalah pemilihan fitur yang nantinya akan diproses dalam metode data mining. Penggunaan sedikit fitur yang mempunyai kondisi diskriminan dua kelas dengan baik tentu akan memberikan kinerja sistem lebih baik daripada penggunaan banyak fitur tetapi tidak memberikan diskriminan dua kelas dengan baik. Dalam penelitian ini, pemilihan fitur dilakukan dengan cara memeriksa ada tidaknya diskriminasi antar kelas dengan pengujian hipotesis *Mann-Whitney*. Ide dasarnya adalah menguji apakah terdapat perbedaan nilai rata-rata fitur setiap kelas secara signifikan.

Hipotesis  $H_0: E(X) = E(X)$

$H_1: E(X) \neq E(X)$

Taraf Signifikansi  $\alpha = 5\% = 0,05$

Statistik Uji

Setelah dilakukan pengolahan menggunakan matlab, diperoleh hasil sebagai berikut:

**Tabel 2.** Pemilihan Fitur Menggunakan Uji Mann-Whitney

Fitur	$Z_{hitung}$	P-value
Kupon	15.7804	$4.2480 \times 10^{-56}$
Jangka Waktu Obligasi	3.2064	0.0013
Peringkat Obligasi	-2.2740	0.0230
Nilai Nominal	-1.1427	0.2532
Yield	-4.5672	$4.9426 \times 10^{-6}$

Keputusan

Berdasarkan statistik uji, dapat dilihat bahwa nilai p-value dari fitur kupon, jangka waktu obligasi, peringkat obligasi dan yield kurang dari  $\alpha$  (0,05), sehingga  $H_0$  ditolak. Sedangkan fitur nilai nominal mempunyai p-value yang lebih dari  $\alpha$  (0,05), oleh karena itu  $H_0$  diterima. Sehingga dapat disimpulkan bahwa fitur kupon, jangka waktu obligasi, peringkat obligasi dan *yield* mempunyai rata-rata yang berbeda signifikan antar kelasnya. Sedangkan fitur nilai nominal mempunyai rata-rata yang sama antar kelasnya. Oleh karena



itu, hanya fitur kupon, jangka waktu obligasi, peringkat obligasi dan *yield* yang digunakan dalam proses klasifikasi.

#### 4.2. Teknik Validasi Model

Teknik validasi yang digunakan dalam penelitian ini adalah *k-fold cross validation*. *K-fold cross validation* memecah data menjadi *k* bagian. Setiap data akan menjadi data uji (*testing*) dan menjadi data latih (*training*) sebanyak *k-1* kali. Penelitian ini menggunakan *k* sebanyak 6. Dengan menggunakan software matlab, setiap data akan diberi kode 1 s.d. 6 secara acak. Ketika data dengan kode *k* = 1 menjadi *testing*, maka data dengan kode *k* = 2, 3, 4, 5, dan 6 akan menjadi *training*. Demikian juga ketika data dengan kode *k* = 2 menjadi *testing*, maka data dengan kode *k* = 1, 3, 4, 5, dan 6 menjadi data *training*. Prosedur ini diulangi sampai data *testing* dengan kode *k* = 6. Pengolahan data menggunakan teknik validasi *k-fold cross validation* dengan *k* = 6, yang diulang sebanyak 30 kali menghasilkan akurasi rata-rata terbaik 78,52%. Berikut contoh perhitungan manual ketika *k* = 1 menjadi data *testing*:

**Tabel 3.** Data *Training* ketika *k* = 1

No	Kupon	Jangka Waktu	Peringkat Obligasi	Nilai Nominal	Yield	Kelas	Indices
1	9,35	2	3	375	10,96829	0	5
2	9,35	2	3	375	10,96858	0	6
3	8,1	2	3	125	11,27969	0	2
4	8,1	2	3	125	11,29308	0	6
5	9,8	2	3	250	11,64721	0	5
603	9,75	2	3	1150	10,92358	0	3

\*nilai nominal dalam milyar rupiah

Dari data *training* tersebut, akan diprediksi data *testing* dengan kupon obligasi sebesar 9,8%, Jangka waktu terletak di kategori 2, peringkat obligasi A, *yield* 11,61629, dan kelas asli data adalah diskon.

Langkah-langkah yang dilakukan untuk memprediksi menggunakan metode *Naive Bayes* adalah sebagai berikut:

1. Hal pertama yang dilakukan adalah menghitung probabilitas prior ( $P(Y)$ ) dari data *testing* berdasarkan data *training*.

**Tabel 4.** Jumlah Data *Training* Per Kelas

Kelas 0 (Diskon)	Kelas 1 (Premium)	Total
350	252	602

Probabilitas prior kelas diskon,  $P(Y = 0) = \frac{350}{602} = 0,5841$

Probabilitas prior kelas premium,  $P(Y = 1) = \frac{252}{602} = 0,4186$

2. Menghitung probabilitas atribut terhadap masing-masing kelas ( $P(X_i|Y)$ ) pada data *testing* berdasarkan data *training*.

1) Kupon Obligasi. Kupon obligasi tidak bertipe kategorik, maka terlebih dahulu dihitung nilai mean dan varian setiap kelasnya untuk menghitung  $P(X|Y)$ .

a) Kelas 0 (Diskon)

$$\text{Mean} = \frac{9,35 + 9,35 + 8,1 \dots + 9,75}{350} = 9,1423$$

$$\text{Varian} = \frac{(9,35 - 9,1423)^2 + \dots + (9,75 - 9,1423)^2}{350 - 1} = 1,2348$$

$$\text{Standar deviasi} = \sqrt{1,2348} = 1,1112$$

$$P(X_1 = 9,8 | 0) = \frac{1}{1,1112\sqrt{2\pi}} e^{-\frac{(9,8-9,1423)^2}{2 \times 1,2348}} = 0,3013$$

b) Kelas 1 (Premium)

$$\text{Mean} = \frac{10,5 + 10,5 + \dots + 12,55}{252} = 10,7984$$

$$\text{Varian} = \frac{(10,5 - 10,7984)^2 + \dots + (12,55 - 10,7984)^2}{252 - 1} = 1,3536$$

$$\text{Standar deviasi} = \sqrt{1,3536} = 1,1635$$

$$P(X_1 = 9,8 | 1) = \frac{1}{1,1635\sqrt{2\pi}} e^{-\frac{(9,8-10,7984)^2}{2 \times 1,3536}} = 0,2373$$

2) Jangka Waktu Obligasi

**Tabel 5.** Jangka Waktu Obligasi pada Data *Training*

Jangka Waktu	Kelas 0	Kelas 1	Total
JW = 2	323	214	537
JW = 3	27	38	65
total	350	252	602

a) Kelas 0 (Diskon)  $P(X_2 = 2 | 0) = \frac{323}{350} = 0,9229$

b) Kelas 1 (Premium)  $P(X_2 = 2 | 1) = \frac{214}{252} = 0,8492$

3) Peringkat Obligasi

**Tabel 6.** Peringkat Obligasi pada Data *Training*

Peringkat	Kelas 0	Kelas 1	Total
1	94	102	196
2	131	67	198
3	111	78	189
4	14	5	19
Total	350	252	602

a) Kelas 0 (Diskon)  $P(X_3 = 3 | 0) = \frac{111}{350} = 0,3171$

b) Kelas 1 (Premium)  $P(X_3 = 3 | 1) = \frac{78}{252} = 0,3095$

4) *Yield*. Dikarenakan *yield* tidak bertipe kategorik, maka terlebih dahulu dihitung nilai mean dan varian setiap kelasnya untuk menghitung  $P(X|Y)$ .

a) Kelas 0 (Diskon)

$$\text{Mean} = \frac{10,96829 + 10,96858 + \dots + 10,92358}{350} = 10,3433$$

$$\text{Varian} = \frac{(10,96829 - 10,3433)^2 + \dots + (10,92358 - 10,3433)^2}{350 - 1} = 1,0540$$

$$\text{Standar deviasi} = \sqrt{1,0540} = 1,0266$$

$$P(X_4 = 11,61629 | 0) = \frac{1}{1,0266\sqrt{2\pi}} e^{-\frac{(11,61629 - 10,3433)^2}{2 \times 1,0540}} = 0,1801$$



b) Kelas 1 (Premium)

$$\text{Mean} = \frac{8,942852 + 8,959918 + \dots + 10,89169}{252} = 9,9924$$

$$\text{Varian} = \frac{(8,942852 - 9,9924)^2 + \dots + (10,89169 - 9,9924)^2}{252 - 1} = 0,9363$$

$$\text{Standar deviasi} = \sqrt{0,9363} = 0,9676$$

$$P(X_4 = 11,61629|1) = \frac{1}{0,9676\sqrt{2\pi}} e^{-\frac{(11,61629 - 9,9924)^2}{2 \times 0,9363}} = 0,1008$$

3. Menghitung perkalian probabilitas prior dengan probabilitas atribut pada masing-masing kelas ( $P(Y) \sum_{i=1}^4 P(X_i|Y)$ ).

1) Data *testing* 1

$$\begin{aligned} \sum_{i=1}^4 P(X_i|Y=0) &= P(X_1=9,8|0) \times P(X_2=2|0) \times P(X_3=3|0) \times P(X_4=11,616|0) \\ &= 0,3013 \times 0,9229 \times 0,3171 \times 0,1801 = 0,0159 \end{aligned}$$

$$\begin{aligned} \sum_{i=1}^4 P(X_i|Y=1) &= P(X_1=9,8|1) \times P(X_2=2|1) \times P(X_3=3|1) \times P(X_4=11,616|1) \\ &= 0,2373 \times 0,8492 \times 0,3095 \times 0,10083 = 0,0063 \end{aligned}$$

$$P(Y=0|X) = P(Y=0) \times \sum_{i=1}^4 P(X_i|Y=0) = 0,0092$$

$$P(Y=1|X) = P(Y=1) \times \sum_{i=1}^4 P(X_i|Y=1) = 0,0026$$

Berdasarkan perhitungan tersebut diperoleh  $P(Y=0|X)$  lebih besar dari  $P(Y=1|X)$  sehingga dapat disimpulkan bahwa data *testing* dengan kupon obligasi sebesar 9,8, jangka waktu antara 1 s.d. 10 tahun atau terletak di kategori 2, peringkat obligasi A atau di kategori 3 dan *yield* sebesar 11,61629 dengan menggunakan metode *Naive Bayes* diperoleh hasil prediksi di kelas 0 (diskon). Hal ini sesuai dengan kelas asli data.

#### 4.3. Pengukuran Kinerja Klasifikasi

*K-fold cross validation* membagi data menjadi  $k$  bagian. Setiap  $k$  bagian mempunyai akurasi dan *error*. Akurasi total diperoleh dengan menambahkan semua jumlah prediksi yang benar dari setiap  $k$  bagian. Sehingga cara untuk mendapatkan akurasi rata-rata yaitu akurasi total dibagi dengan  $k$ .

$$\text{Akurasi rata-rata} = \frac{\text{jumlah prediksi yang benar}}{\text{jumlah prediksi keseluruhan}} = \frac{94 + 106 + 85 + 88 + 94 + 103}{726} = 0,7852$$

$$\text{Error} = 1 - 0,7852 = 0,2148$$

Pengklasifikasian data perubahan harga obligasi korporasi di Indonesia menggunakan metode *Naive Bayes* dengan teknik validasi *k-fold cross validation* dengan  $k = 6$  diperoleh akurasi rata-rata sebesar 78,52% dan *error* 21,48%. Dengan akurasi yang cukup tinggi dan *error* yang cukup rendah, hal ini dapat diartikan bahwa metode *Naive Bayes* bekerja cukup baik dalam mengklasifikasi perubahan harga obligasi korporasi di Indonesia.

## 5. PENUTUP

### 5.1. Kesimpulan

Berdasarkan hasil dan pembahasan diperoleh kesimpulan bahwa:

1. Variabel independen yang digunakan dalam penelitian ini berjumlah 5, yaitu kupon obligasi, jangka waktu obligasi, peringkat obligasi, nilai nominal, dan *yield*. Akan tetapi, setelah dilakukan *preprocessing data* menggunakan uji *Mann-Whitney* yang bertujuan untuk mendapatkan fitur yang independen tiap kelasnya, diperoleh kesimpulan bahwa nilai nominal tidak mempunyai diskriminasi yang kuat antar kelasnya. Sehingga hanya kupon obligasi, jangka waktu obligasi, peringkat obligasi, dan *yield* yang digunakan dalam proses pengklasifikasian.

2. Dengan teknik validasi *k-fold cross validation* dengan  $k = 6$ , diperoleh akurasi rata-rata sebesar 78,52% dan *error* 21,48%. Hal ini dapat diartikan bahwa metode *Naive Bayes* bekerja cukup baik untuk mengklasifikasi perubahan harga obligasi korporasi di Indonesia tanggal 30 November 2015 sampai dengan 1 Desember 2015.

## 5.2. Saran

Metode *Naive Bayes* merupakan metode yang cukup baik untuk mengklasifikasikan perubahan harga obligasi korporasi di Indonesia. Harga obligasi yang mengalami perubahan setiap waktu dapat menguntungkan atau merugikan investor. Dengan mengklasifikasikan perubahan harga obligasi dapat membantu investor untuk mendapatkan *return* yang optimal. Apabila hasil prediksi dari proses klasifikasi menunjukkan kelas 0 maka kondisi terbaik untuk investor adalah membeli obligasi dengan harga diskon. Apabila hasil prediksi dari proses klasifikasi menunjukkan kelas 1 maka kondisi terbaik untuk investor adalah menjual obligasi dengan harga premium.

## DAFTAR PUSTAKA

- Ahmad, K. 2003. *Dasar-dasar Manajemen Investasi dan Portofolio*. Jakarta: Rineka Cipta.
- Anoraga, P. dan Pakarti, P. 2001. *Pengantar Pasar Modal*. Jakarta: Rineka Cipta.
- Daniel, W. W. 1989. *Statistika Nonparametrik Terapan*. Alex Tri K. W., penerjemah. Jakarta: Gramedia. Terjemahan dari *Applied Nonparametric Statistics*.
- Fabozzi, F. J. 2000. *Manajemen Investasi Buku 2*. Jakarta: Penerbit Salemba Empat.
- Han, J dan Kamber, M. 2006. *Data Mining Concepts and Techniques, Second Edition*. California: Morgan Kaufman.
- Ibrahim, H. 2008. *Pengaruh Tingkat Suku Bunga, Peringkat Obligasi, Ukuran Perusahaan dan DER Terhadap Yield to Maturity Obligasi Korporasi di Bursa Efek Indonesia Periode Tahun 2004-2006*. Tesis Magister Manajemen. Pasca Sarjana, Universitas Diponegoro.
- Krisnilasari, M. 2007. *Analisis Pengaruh Likuiditas Obligasi, Coupon dan Jangka Waktu Jatuh Tempo Obligasi Terhadap Harga Obligasi di Bursa Efek Surabaya*. Tesis Magister Manajemen. Pasca Sarjana, Universitas Diponegoro.
- Prasetyo, E. 2013. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Penerbit Andi.
- Prasetyo, E. 2014. *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: Penerbit Andi.
- Santosa, B. 2007. *Data mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sibarani, Y. 2008. *Analisis dan Penerapan Metode Klasifikasi untuk Membangun Perangkat Lunak Sistem Penerimaan Mahasiswa Baru Jalur Non Tulis*. Tesis Magister Informatika. Institut Teknologi Bandung.
- Subanar. 2013. *Statistika Matematika: Probabilitas, Distribusi, dan Asimtotis dalam Statistika*. Yogyakarta: Graha Ilmu.
- Wijayanto, S. 2015. *Strategi Jitu Investasi Obligasi Korporasi: Sudut Pandang Investasi dari Seorang Analis*. Jakarta : Penerbit PT. Elex Media Komputindo.
- Zubir, Z. 2012. *Portofolio Obligasi*. Jakarta: Penerbit Salemba Empat.